DOCUMENT RESUME

ED 395 935 TM 024 990

AUTHOR	Tucker, Ledyard R.
TITLE	A Continuity Principle for Calibration of Scores within Mastery Assessment Systems.
INSTITUTION	Educational Testing Service, Princeton, N.J.
REPORT NO	ETS-RR-91-53
PUB DATE	Mar 92
NOTE	60p.
PUB TYPE	Reports - Evaluative/Feasibility (142)
EDRS PRICE	MF01/PC03 Plus Postage.
DESCRIPTORS	Computation; *Educational Assessment; *Mastery Learning; Performance Factors; *Scaling; *Scores; Training
IDENTIFIERS	*Calibration; *Continuity; Curves

ABSTRACT

A continuity principle is suggested for scaling of assessment scores from different levels of a multilevel mastery training program. In such training programs students are self-paced and work at levels of tasks appropriate for their levels of performance. The problem addressed in this report concerns the scaling of assessment scores at different levels so as to form a single basic scale. The solution proposed here involves relative scalings of the scores at different levels so that the mean converted score over all students forms a smooth, continuous curve. The smoothness of this mean curve is measured by higher order differences between consecutive scores. Computational procedures to obtain as smooth a curve as possible are presented. (Contains 11 tables, 9 figures, technical notes, and 1 reference.) (Author)

****

* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *

****

ED 395 935

RESEARCH REPORT

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
This document has been reproduced as received from the person or organization originating it
Minor changes have been made to improve reproduction quality
• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy

PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY
H. I. Braun
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

A CONTINUITY PRINCIPLE FOR CALIBRATION OF SCORES WITHIN MASTERY ASSESSMENT SYSTEMS

Ledyard R Tucker

Educational Testing Service
Princeton, New Jersey
March 1992

BEST COPY AVAILABLE

TM024990

A CONTINUITY PRINCIPLE FOR CALIBRATION OF SCORES WITHIN MASTERY ASSESSMENT SYSTEMS

Ledyard R Tucker

Educational Testing Service
Princeton, New Jersey 08541

ERIC

Copyright © 1992. Educational Testing Service. All rights reserved.

A CONTINUITY PRINCIPLE FOR CALIBRATION OF SCORES

WITHIN MASTERY ASSESSMENT SYSTEMS

Ledyard R Tucker

ABSTRACT

A continuity principle is suggested for scaling of assessment scores from different levels of a multilevel mastery training program. In such training programs students are self-paced and work at levels of tasks appropriate for their levels of performance. The problem addressed in this report concerns the scaling of assessment scores at different levels so as to form a single basic scale. The solution proposed here invloves relative scalings of the scores at different levels so that the mean converted score over all students forms a smooth, continuous curve. The smoothness of this mean curve is measured by higher order differences between consecutive scores. Computational procedures to obtain as smooth a curve as possible are presented.

A CONTINUITY PRINCIPLE FOR CALIBRATION OF SCORES WITHIN MASTERY ASSESSMENT SYSTEMS*

Ledyard R Tucker

This report concerns scaling of assessment scores from multilevel mastery training programs in specified areas of study. Students in such a program may be working in a self-paced manner on tasks appropriate to the students' levels of progress. A number of diagnostic observations may be made during students' work on tasks with feedback being given. In addition, assessments may be made of students' work during the time of their performance. Students advance from one level to another based upon their assessment scores. Students could be set back to previous task levels based on poor assessment scores. There is a problem of scaling these assessment scores from task level to task level to form a combined scale. One suggestion to accomplish this scaling involves a principle of continuity from level to level.

*Discussions of the problems with Peter Pashley and Charles Lewis were extremely helpful. Their contributions are gratefully acknowledged.

Introduction

A training program is conceived, here, as composed of a series of sessions which may be divided into subsessions during which each student works on a task. When student activity is self paced, different students may work on different tasks during a session. The unit of activity is termed, here, as a subsession with records being kept by subsession. A subsession may be considered analogous to a trial in an experiment. (The term trial is used in simulation, Monte Carlo studies to be described later.) A major score and, possibly, several diagnostic scores are derived from the behavior of a student in performance of a task. These scores need not be derived from tests composed of items. For example in a psycho-motor experiment involving a tracking task, time on target could be a score for each trial. These scores are reflections of qualities of performances on tasks. Each type of score is taken to be a complex composite of skills which are being trained. Each skill may be dependent, in part, on several latent traits, both abilities and personality traits, with the nature of this dependency shifting as the training progresses. Further, the levels, or scores, on these latent traits may change as influenced by the training program. Thus, each type of score is quite complex over a series of subsessions. This poses considerable problems in linking scores over different tasks, especially tasks at different levels of difficulty and complexity.

There may be a collection of more or less equivalent tasks at a given level so that students working at this level do not repeat a fixed task for that level. Scores on these tasks may be equated by any of several experimental methods. One such method would involve dividing, randomly, a sample of individuals with equivalent backgrounds into subsamples. Each subsample would work on one of the tasks at a defined level. Scores of the students in these subsamples could be used to equate the scores for the several tasks. In this approach, scores on other common variables could be used as covariates to reduce sampling error.

Major consideration was given in the present study to linking score scales on tasks at different levels. The complexities referred to in a preceding paragraph along with the need to consider task scores in general eliminate presently available calibration procedures. Only a principle which we may term score scale continuity appears to remain. By this term we mean to include smoothness of function of scaled scores of individual performances as related to extent of work on the learning topic. With this principle there should be no major discontinuities, either jumps in score values or drops associated with going from a lower level task to a higher level task. Further, there should be no drastic changes in slope of performance curves at junctions between task levels; general trends should be maintained. There is no necessary requirement that the trend be increasing even though this would be desirable. The major conception is the smoothness of the function of performance on extent of practice.

Investigation of the applicability of this principle was accomplished with a series of simulation, Monte Carlo studies. In these studies a plan for implementation was developed along with tentative computational procedures. Four types of simulation systems were used ranging from a very simple situation involving only a general learning function to increased complexity of learning situations. Discussion of simulation system type 1 will present the general pattern of operations. Simulation system type 2 introduces errors of measurement and a system for individual review and trials at the advanced level. Simulation system type 3 introduces specific skills for each of two levels of tasks. Simulation type 4 introduces differences in learning functions for different simulated individuals. There were early learners and late learners. Each simulation system type will be discussed in a separate section.

Simulation System Type 1

In the present simulation system, type 1, a single general latent trait was conceived on which individual scores increased with practice. There were two levels of tasks with the second level, in comparison with the first level of task, being more difficult and having an observed score for which the unit of measure was smaller (as a consequence, the standard deviation was larger). A series of 15 trials was employed with each simulated individual working on a task at each trial. Performance of a simulated individual on a task yielded an observed score. Each such individual started on task level 1 and continued at this level until its observed score equalled or exceeded a cutting score at which time it was advanced to task level 2. If a simulated individual's performance on task level 2 was below a cutting score for that task, it was returned to task level 1. In simulation system type 1 all switching between task levels occurred as described above. In subsequent simulation system types additional switching between task levels was enforced to provide either a look ahead from task level 1 to task level 2 or a review of task level 1 for an individual working at task level 2. This additional task level switching did not take place in simulation system type 1. A result for each simulated individual was a series of scores, some on task level 1 and some on task level 2. The object of the analysis was to provide a conversion of task level 2 scores such that a combined series of scores, task level 1 scores and converted task level 2 scores, appeared to be on a single scale. A linear conversion was employed for this purpose.

TABLE 1

PARAMETERS FOR SIMULATION SYSTEM TYPE 1

RANDOM SCORE RANGES

Initial Score	5 to 25
Asymptote	35 to 55

TASK LEVEL 2 SCALING COEFFICIENTS

Additive	-20.0
Multiplicative	1.25

CUTTING SCORES

Task Level 1:	30 to Advance to Task Level 2
Task Level 2:	27 to Remain at Task Level 2

SAMPLE SIZE: 1000

Tables 1, 2, 3 and 4 pertain to simulation system type 1, run 1. Parameters for this run are given in Table 1. Two, independent, random scores were drawn for each simulated individual; the trait score at trial 1 and the asymptote trait score. Each of these scores was drawn from a triangular distribution having the ranges given in Table 1. Trait scores on the trials were generated by a two parameter, negative exponential learning function. Level 1 observed scores equalled the trait scores. For task level 2, the observed scores were a linear transformation of the trait scores for the trials, this transformation using the listed scaling coefficients. There were two cutting scores: a task level 1 score to advance to task level 2 and a task level 2 score to remain at task level 2. A sample of 1000 simulated individuals was run.

TABLE 2

EXAMPLES OF INDIVIDUAL SCORES

SIMULATION SYSTEM TYPE 1

Individual	Task level	Trial 1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1	1	20	30	--	38	--	--	--	--	--	--	--	--	--	--	--
	2	--	--	24	--	29	30	30	31	31	31	31	31	31	31	31
	C*	20	30	35	38	39	40	40	41	41	41	41	41	41	41	41
2	1	12	20	27	32	--	39	--	--	--	--	--	--	--	--	--
	2	--	--	--	--	25	--	31	33	35	36	37	38	38	39	39
	C*	12	20	27	32	36	39	41	43	44	45	46	47	47	47	47
3	1	13	21	26	29	31	--	33	--	33	--	33	--	34	--	34
	2	--	--	--	--	--	20	--	21	--	22	--	22	--	22	--
	C*	13	21	26	29	31	32	33	32	33	34	33	34	34	34	34
4	1	18	29	34	--	--	--	--	--	--	--	--	--	--	--	--
	2	--	--	--	27	30	31	32	33	33	33	33	33	33	33	33
	C*	18	29	34	38	40	41	42	43	43	43	43	43	43	43	43

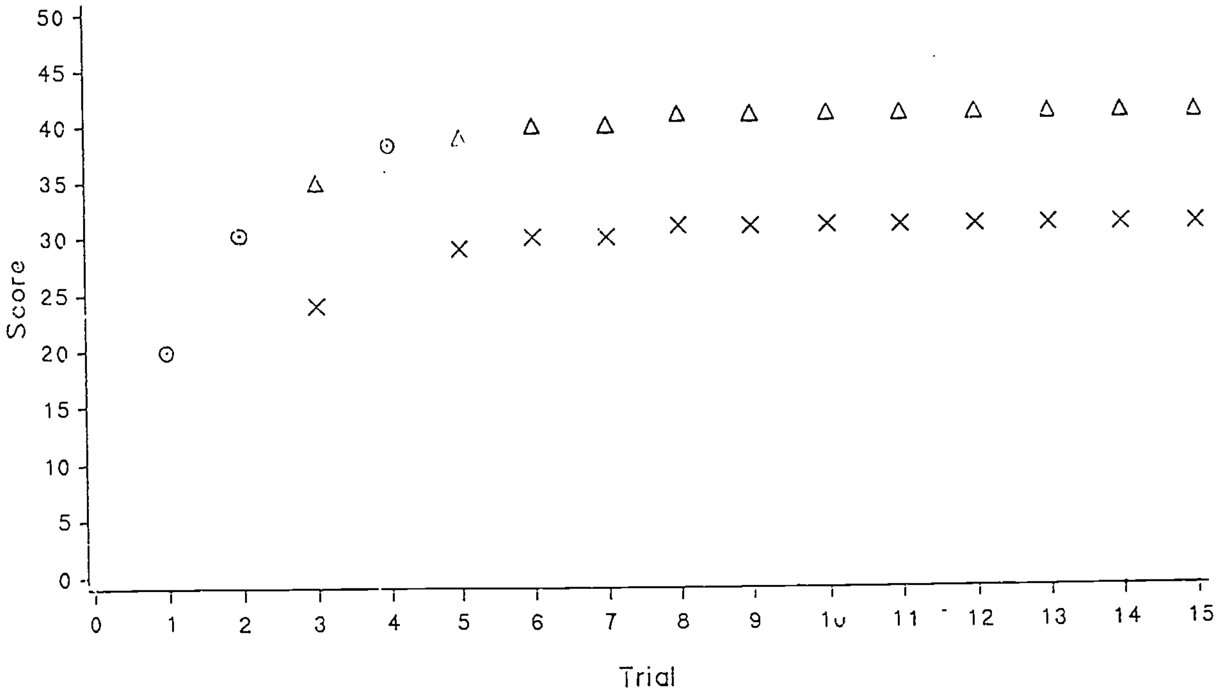
*Task level C contains task level 1 scores and converted task level 2 scores; conversion coefficients: a = 15.5; b = .82.

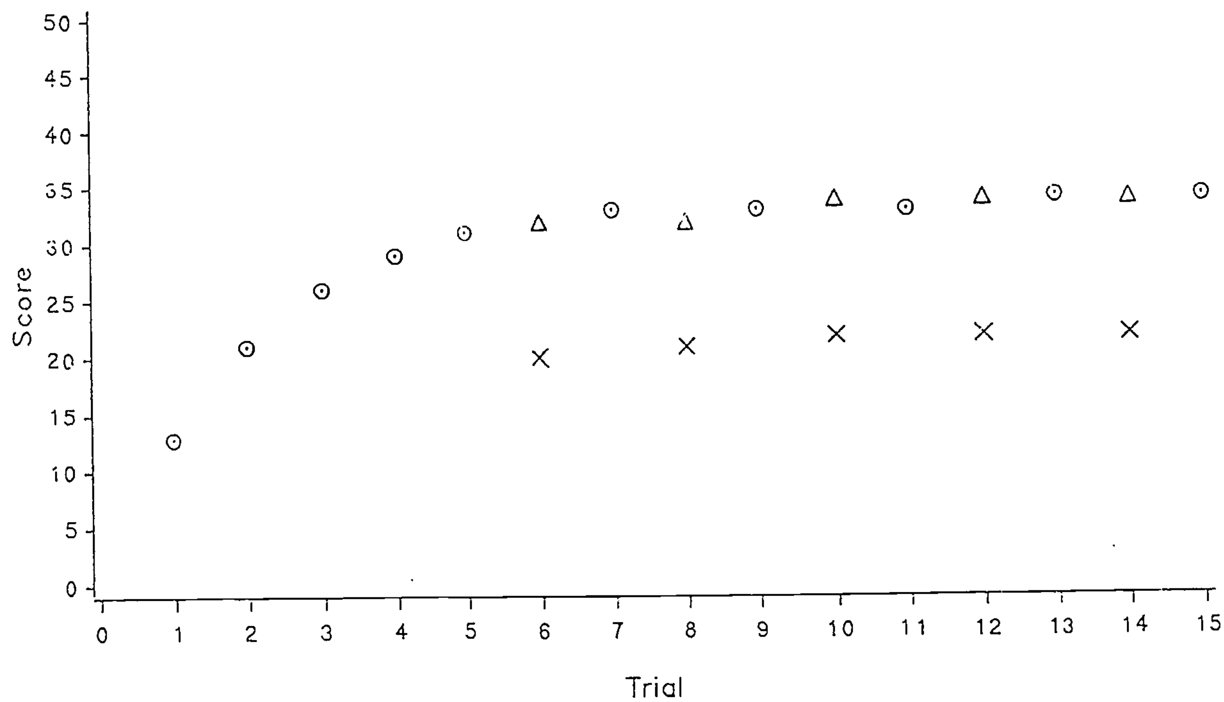
Scores for four simulated individuals are given in Table 2. Look at rows for task levels 1 and 2; ignore, for the present, rows labeled C.

Individual 1 started on task level 1 and by trial 2 had a score equal to the task level 1 cutting score; consequently this individual was advanced to task level 2 for trial 3. However, this individual's level 2 score on trial 3 was less than the cutting score to remain at task level 2; therefore, this individual was returned to task level 1 for trial 4. Its trial 4 score again exceeded the task level 1 cutting score so it was advanced to task level 2 for trial 5. This individual remained at task level 2 for the remaining trials. This individual's performance is shown in the upper graph of Figure 1. Look at only the circles for observed task level 1 scores and x's for observed task level 2 scores. Ignore, for the present, the triangles.

Simulated individual 2 had a similar record to simulated individual 1 except that it took four trials before the first switch to task level 2. Simulated individual 3 had a different kind of record in that this individual alternated between the two task levels. Evidently this individual's asymptote was such that it could perform task level 1 satisfactorily but not task level 2. This individual's performance is shown in the lower graph of Figure 1. Simulated individual 4 is an illustration of an individual which did not switch back to task level 1 once having achieved task level 2.

A conversion for task level 2 scores was computed by a procedure to be described in a subsequent paragraph. Coefficients for this linear conversion are given at the bottom of Table 2; a is an additive coefficient and b is a multiplicative coefficient. A perfect inverse transformation





⊙ ⊙ ⊙ Original Scores; Task Level 1
× × × Original Scores; Task Level 2
Δ Δ Δ Converted Scores; Task Level 2

Figure 1. Examples of individual performances; simulation system, type 1.

from the scaling coefficients would have a = 16 and b = .8 . The values given in Table 2 were obtained from the sample data. Row C scores in Table 2 for each simulated individual have task level 1 scores when the individual worked on this task and converted task level 2 scores when the individual worked at task level 2. The scores in row C form the combined score vector for the simulated individual. In Figure 1 the converted task level 2 scores are indicated by triangles so that each combined score vector is a series of circles and triangles. Note that the combined score vectors are quite continuous.

Various experimental conversion coefficients may be tried to obtain corresponding experimental combined score vectors. One such vector for each individual could be recorded in an experimental combined score matrix having a row for each individual and a column for each trial. A mean combined score vector could be obtained from this matrix, this vector would have the mean score for each column, or trial. The top half of Table 3 presents information for experimental conversion coefficients a = 0.0 and b = 1.0 , this being equivalent to applying no conversion to task level 2 scores, that is, using the original task level 2 scores. Figure 2 presents the series of mean scores (x's connected by lines) when the original task level 2 scores are used.

One way to look for continuity in such a vector is to examine differences between successive mean scores. Level of differences 1 are the mean gains from trial to trial; for example, trial 2 mean of 23.9 minus trial 1 mean of 15.0 is 8.9 which is recorded as level of differences 1 in the row for trial 2. The 4.0 in row 3 is the mean gain from trial 2 to trial 3. The remaining entries in level of differences 1 are the

BEST COPY AVAILABLE

TABLE 3

COMBINED TRIAL MEANS AND DIFFERENCES*
SIMULATION SYSTEM TYPE 1, RUN 1

Level 2 Conversion Coefficients: a = 0.0; b = 1.00

Trial	Mean Score	Level of Differences 1	2	3	4	5
1	15.0					
2	23.9	8.9				
3	27.9	4.0	-4.9			
4	28.5	.6	-3.4	1.5		
5	30.1	1.7	1.0	4.4	3.0	
6	30.2	.1	-1.6	-2.6	-7.1	-10.0
7	30.3	.1	.0	1.6	4.2	11.3
8	30.7	.4	.3	.3	-1.3	-5.5
9	31.0	.3	-.1	-.4	-.7	.7
10	31.0	.0	-.3	-.2	.2	.8
11	31.2	.3	.3	.7	.9	.7
12	31.2	-.1.	-.4	-.7	-1.3	-2.2
13	31.5	.3	.4	.8	1.4	2.8
14	31.2	-.2	-.6	-1.0	-1.7	-3.1
15	31.6	.3	.6	1.1	2.1	3.8
RMSD		2.7	1.8	1.7	2.9	5.4

Level 2 Conversion Coefficients: a = 15.5; b = .82

Trial	Mean Score	Level of Differences 1	2	3	4	5
1	15.0					
2	23.9	8.9				
3	29.3	5.5	-3.4			
4	32.8	3.4	-2.0	1.4		
5	35.0	2.2	-1.2	.8	-.6	
6	36.5	1.5	-.7	.3	-.3	.3
7	37.5	1.0	-.5	.3	-.2	.1
8	38.2	.7	-.3	.2	-.1	.1
9	38.7	.5	-.2	.1	-.1	.0
10	39.0	.4	-.1	.1	-.1	.0
11	39.3	.3	-.1	.1	.0	.0
12	39.5	.2	-.1	.1	.0	.0
13	39.6	.1	.0	.0	.0	.0
14	39.7	.1	.0	.0	.0	.0
15	39.8	.1	.0	.0	.0	.0
RMSD		3.1	1.2	.5	.2	.1

*Inconsistencies in the last digits are due to rounding.

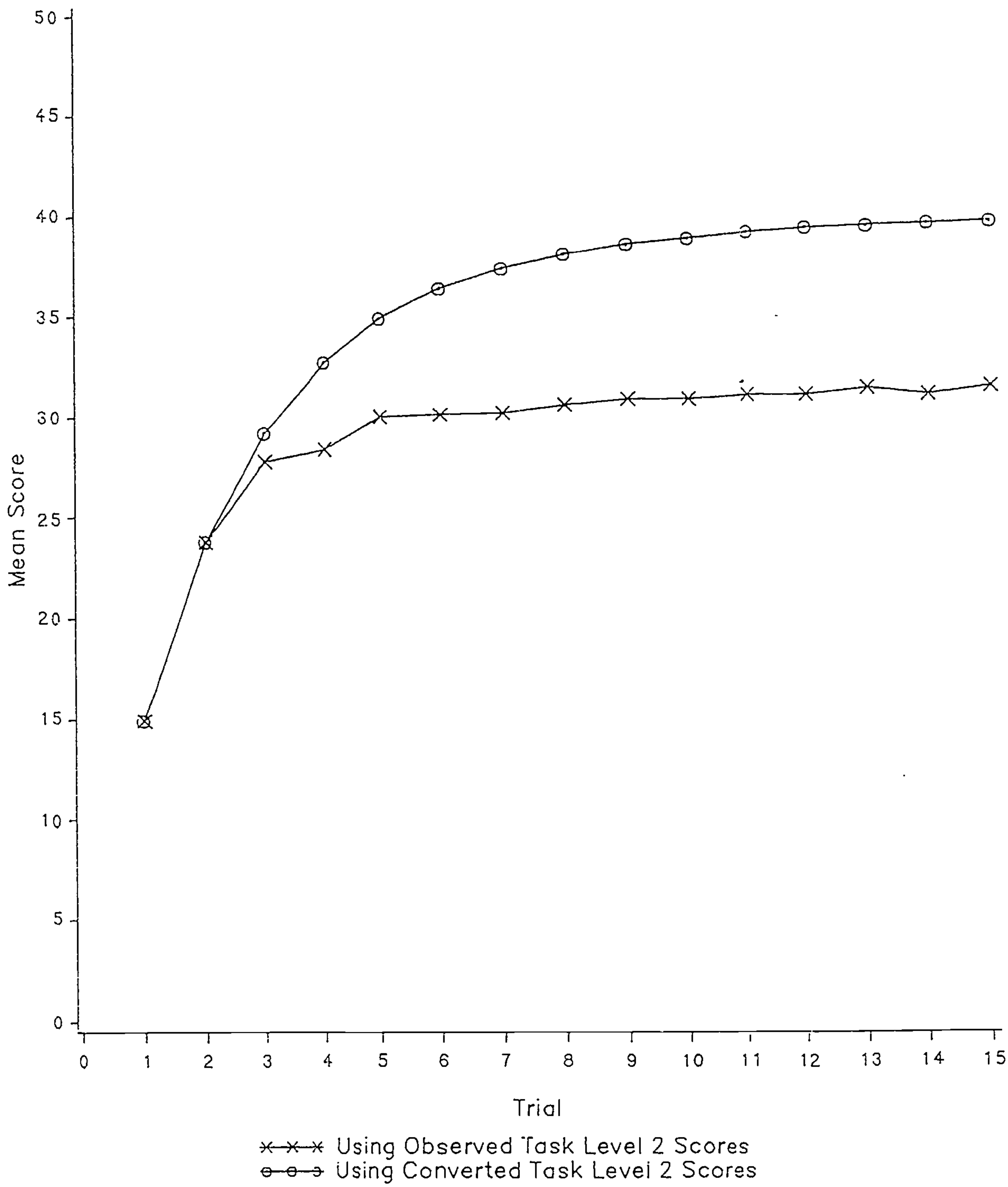


Figure 2. Mean scores for simulation system, type 1, run 1; conversion coefficients: a = 15.5, b = .82.

differences between consecutive trial means. Level of differences 2 are obtained from level of differences 1. Each entry in level of differences 2 is the difference between consecutive entries in level of differences 1. Level of differences 2 are changes in level of differences 1. A similar process of differences of differences may be continued to obtain further levels of differences, each level of differences being obtained from the just preceding level of differences. Thus, level of differences 3 entries are computed from level of differences 2 entries. In Table 3 this process was carried out to level of differences 5. Note some irregularities in each column of level of differences. This is taken as evidence of lack of continuity in the vector of mean combined scores.

At the bottom section of Table 3, results are given using experimental conversion coefficients obtained by a procedure to be described subsequently. Not only are the columns for levels of differences more regular than are the columns in the top half of the table, but also, after level of differences 1, the values are smaller in the lower half than in the upper half of the table. This is especially true for level of differences 5. A measure of the size of entries in each column is obtained with an RMSD statistic. This statistic is obtained by taking the sum of squares of the values in a column, dividing this sum by the number of values in the column and taking the square root. At the bottom of each section of the table is a row of RMSD statistics. For each pair of experimental conversion coefficients and level of differences there would be an RMSD statistic. Note in Table 3 the RMSD coefficients, after level of differences 1, are smaller in the lower half than in the upper half of the table. Continuity of the mean vector could be indicated by small RMSD

ERIC

coefficients in selected levels of differences. Thus, a solution to the conversion problem could be to find the experimental conversion coefficients which yield a minimum RMSD. There remains a problem in selection of the level of differences to be considered. For simulation system type 1, run 1 the level of differences 5 was selected since its minimum RMSD was less than for preceding levels of differences.

Figure 2 presents the two mean vectors listed in Table 3. Pictorially, both series of mean val es appear fairly smooth with the one using converted task level 2 scores appearing slightly smoother at trials 3-5. Use of the levels of differences picked out several irregularities. A method for solution for minimum RMSD at a given level of differences was developed and a computer program was written. Table 4 presents for each

TABLE 4

CONVERSION COEFFICIENTS FOR MINIMUM RMSD

AT LEVELS OF DIFFERENCES 1 THROUGH 5

SIMULATION SYSTEM TYPE 1, RUN 1

Level of Differences	Coefficients a	b	RMSD
1	-11.56	1.181	2.58
2	-2.98	1.767	1.07
3	16.67	.803	.4
4	14.93	.840	.22
5	15.50	.817	.10

level of differences 1 through 5 the conversion coefficients for minimum RMSD for simulation system type 1, run 1. Results for levels of differences 1 and 2 are quite distinct from the results for the higher levels of differences. For these first two levels of differences there are negative additive constants which appear to be unreasonable for an acceptable solution. However, the RMSD statistics for these two levels of differences are somewhat higher than the RMSD statistics for the subsequent levels of differences. In the simple case of this type of simulation a theoretic conversion may be derived from the scaling coefficients used in generation of the data. This is an inverse transformation from the scaling transformation and has conversion coefficients $a = 16$ and $b = .8$. The experimental conversion coefficients at levels of differences 3 through 5 approximate these theoretic coefficients. Level of differences 5 was selected to yield solution conversion coefficients since the RMSD was least at this level of differences.

Simulation System Type 2

Simulation system type 1 was exceedingly simple. Two features were added to produce simulation system type 2: errors of measurement were introduced and a plan for review and look ahead was implemented. In simulation system type 1, scores on the two task levels were compared only at the margin when individuals switched from a task at one level to a task at the other level. There was no opportunity to compare scores on the two levels of tasks at either lower scores or at higher scores. Charles Lewis suggested that score comparisons could be strengthened with observations at both the lower scores and higher scores ranges. This is accomplished with the plan for review and look ahead. Table 5 gives parameters for simulation system type 2. This table may be compared with Table 1. Measurement errors with a range from -1 to +1 were inserted. The other score ranges were not changed and the task level 2 scaling coefficients were not altered. There was some adjustment in the cutting scores. A

TABLE 5

PARAMETERS FOR SIMULATION SYSTEM TYPE 2

RANDOM SCORE RANGES

Initial Score	5 to 25
Asymptote	35 to 55
Measurement Error	-1 to +1

TASK LEVEL 2 SCALING COEFFICIENTS

Additive	-20.0
Multiplicative	1.25

CUTTING SCORES

Task Level 1: 32.5 to Advance to Task Level 2

Task Level 2: 25.0 to Remain at Task Level 2

MAXIMUM NUMBER OF CONSECUTIVE TRIALS:4

SAMPLE SIZE: 1000

maximum number of consecutive trials at a given level of task was introduced to implement the review and look ahead plan. The sample size was kept at 1000.

Examples of simulated individuals' trial scores are given in Figure 3. The individual in the upper graph achieved the task level 1 cutting score at trial 3 and was advanced to task level 2 at trial 4. Its performance continued above the task level 2 cutting score to return to task level 1. After four trials at task level 2, this individual was given a review on task level 1 at trial 8 where its performance was sufficiently high to be advanced again to task level 2. Another review of task level 1 was given at trial 13. These reviews provide score comparisons at a higher score range. The simulated individual's performance illustrated in the lower graph is an example of a lesser ability individual which alternated between the two task levels.

Figure 4 shows two combined mean score vectors: one using the original task level 2 scores and the other involving converted task level 2 scores. In comparison with the combined mean vector graph in Figure 1 for the mean vector using original task level 2 scores, the corresponding mean vector for simulation system type 2 shown in Figure 4 is much more irregular. A possible cause is the adjustment of the cutting scores. Again, however, the conversion of the task level 2 scores provided a very smooth mean vector as shown in Figure 4.

TABLE 6

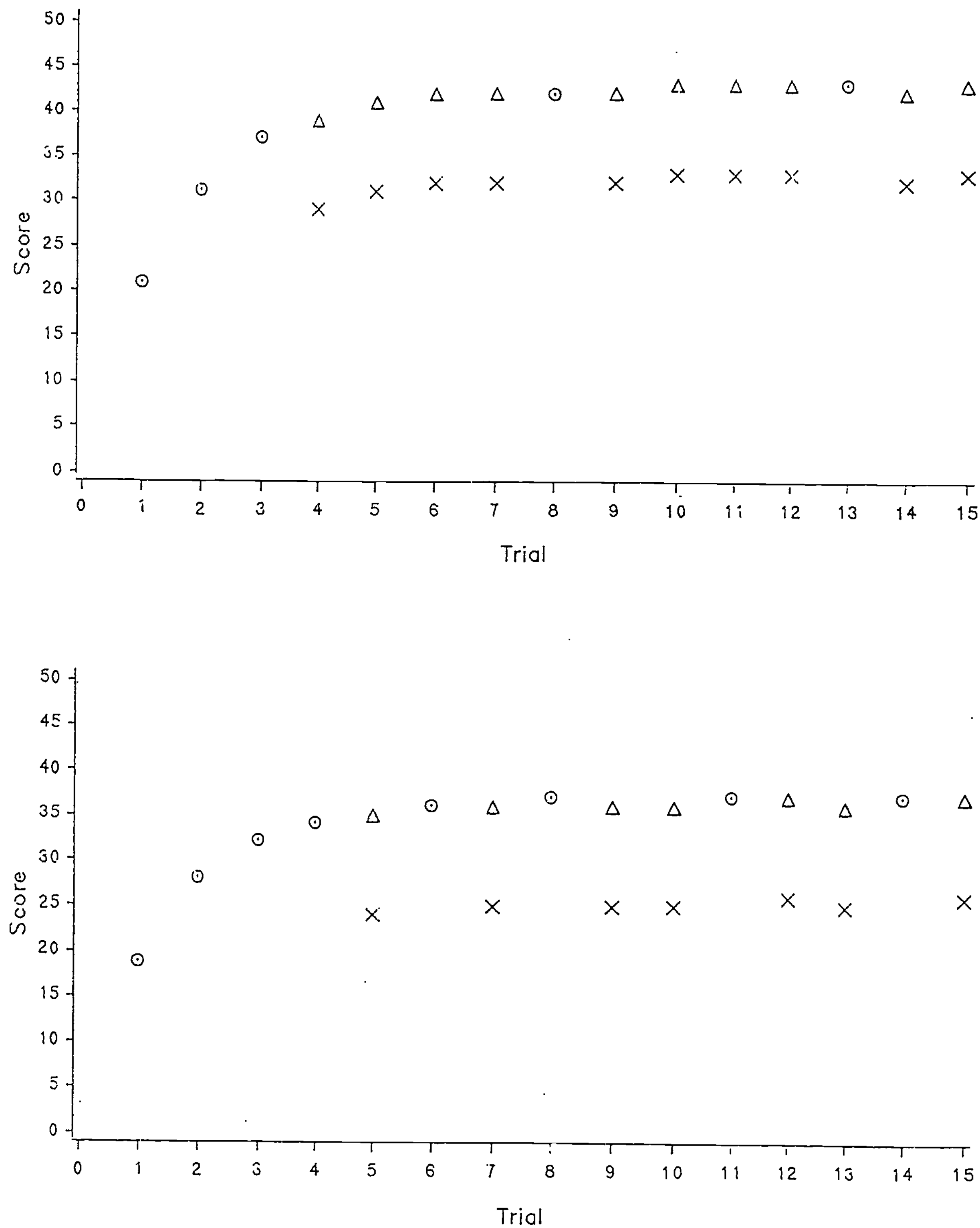
CONVERSION COEFFICIENTS FOR MINIMUM RMSD

AT LEVELS OF DIFFERENCES 1 THROUGH 5

SIMULATION SYSTEM TYPE 2, RUN 1

Level of	Coefficients		
Differences	a	b	RMSD
1	17.51	.642	3.31
2	15.83	.814	1.18
3	16.09	.796	.45
4	16.10	.795	.19
5	16.11	.795	.13

ERIC



ʘ ʘ ʘ Original Scores; Task Level 1
× × × Original Scores; Task Level 2
Δ Δ Δ Converted Scores; Task Level 2

Figure 3. Examples of individual performances; simulation system, type 2.

ERIC

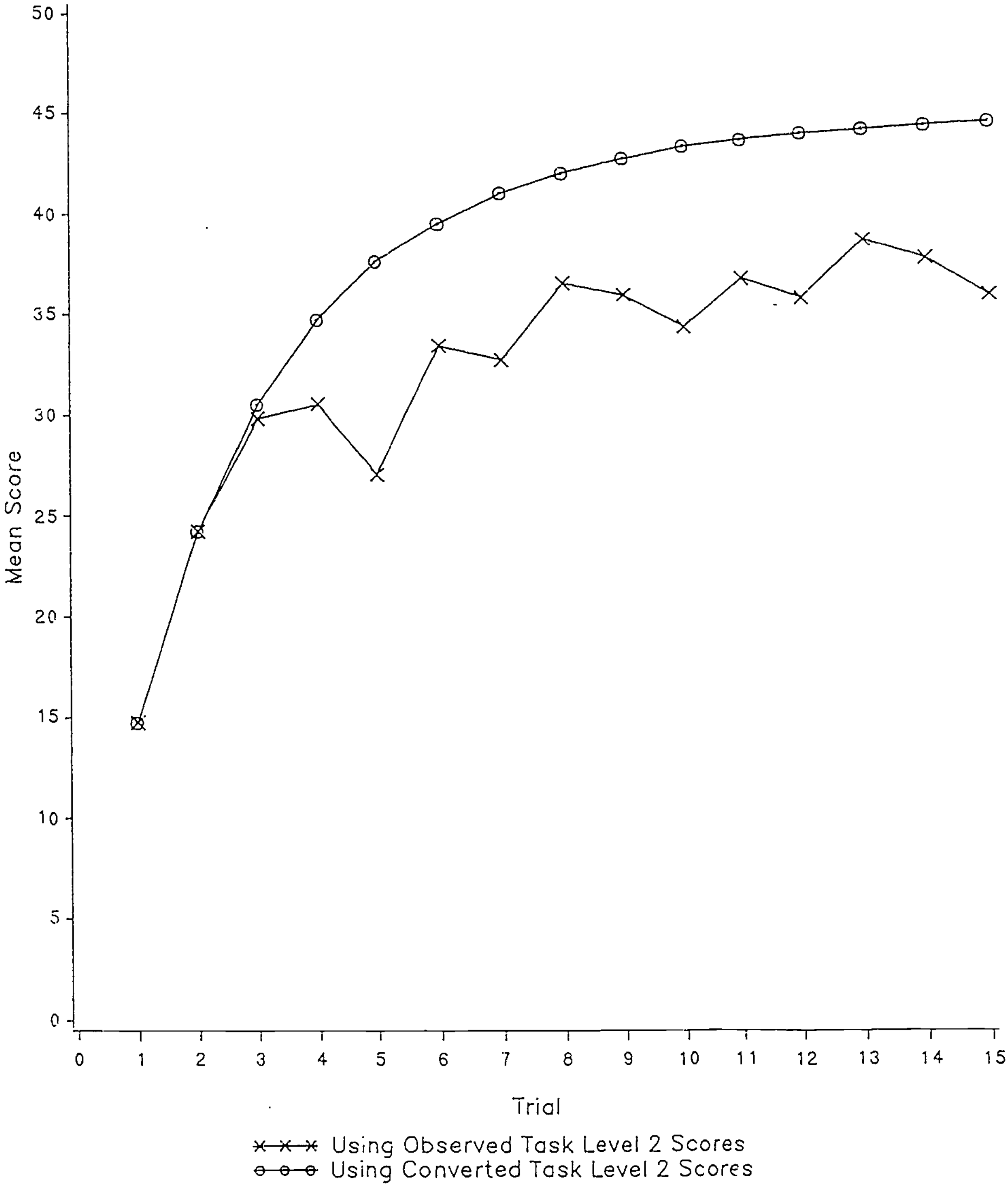


Figure 4. Mean scores for simulation system, type 2, run 1; conversion coefficients: a = 16.1, b = .80.

ERIC

The record of conversion coefficients for minimum RMSD at levels of differences 1 through 5 is given in Table 6. The computed conversion coefficients stabilize from level of differences 2 onward with the least RMSD occurring at level of differences 5. For this example of a simulation system the computed conversion coefficients are much closer to the theoretically derived coefficients of a = 16 and b = .8 than for the example of simulation system type 1. Sampling studies have not been carried out; however, a conjecture is possible that the sampling variance of the conversion coefficients would be much greater for simulation system type 1 than for simulation system type 2. This may be the result of the more extensive score comparisons at higher score ranges with the review and look ahead program in simulation system type 2.

Simulation System Type 3

A feature added in simulation system type 3 was the inclusion of specific traits for each of the task levels. The true scores on each task level were a combination of a general trait and a specific trait. There was a separate specific trait for each task level. Dependence on the specific trait dwindled as an individual had experience at the task level. Experience at one task level did not affect the dependence of the other task level on its specific. Thus, there was a shifting of dependence on the general trait and the specific traits. Growth on the general trait occurred at every trial while growth on each specific trait occurred only on the trials involving the particular task level. Parameters used in simulation system type 3 are given in Table 7. There are separate score ranges for the general trait and the specific traits. Otherwise, the parameters for this simulation system type are the same as for simulation system type 2 with exception of adjustments in the cutting scores.

Examples of performances of two simulated individuals are given in Figure 5. A particular point is that the specific traits started out quite low so that scores on early trials were reduced due to the dependence on the specific traits. As the dependence on the specific traits decreased, this decrement reduced. Note that the observed scores on task level 2 are considerably below the scores on task level 1 in early trials and that this

TABLE 7

PARAMETERS FOR SIMULATION SYSTEM TYPE 3

RANDOM SCORE RANGES

General Initial Score	5 to 25
General Asymptote	35 to 55
Specific Initial Score	1 to 16
Specific Asymptote	35 to 55
Measurement Error	-1 to +1

TASK LEVEL 2 SCALING COEFFICIENTS

Additive	-20.0
Multiplicative	1.25

CUTTING SCORES

Task Level 1: 32.0 to Advance to Task Level 2

Task Level 2: 20.0 to Remain at Task Level 2

MAXIMUM NUMBER OF CONSECUTIVE TRIALS: 4

SAMPLE SIZE: 1000

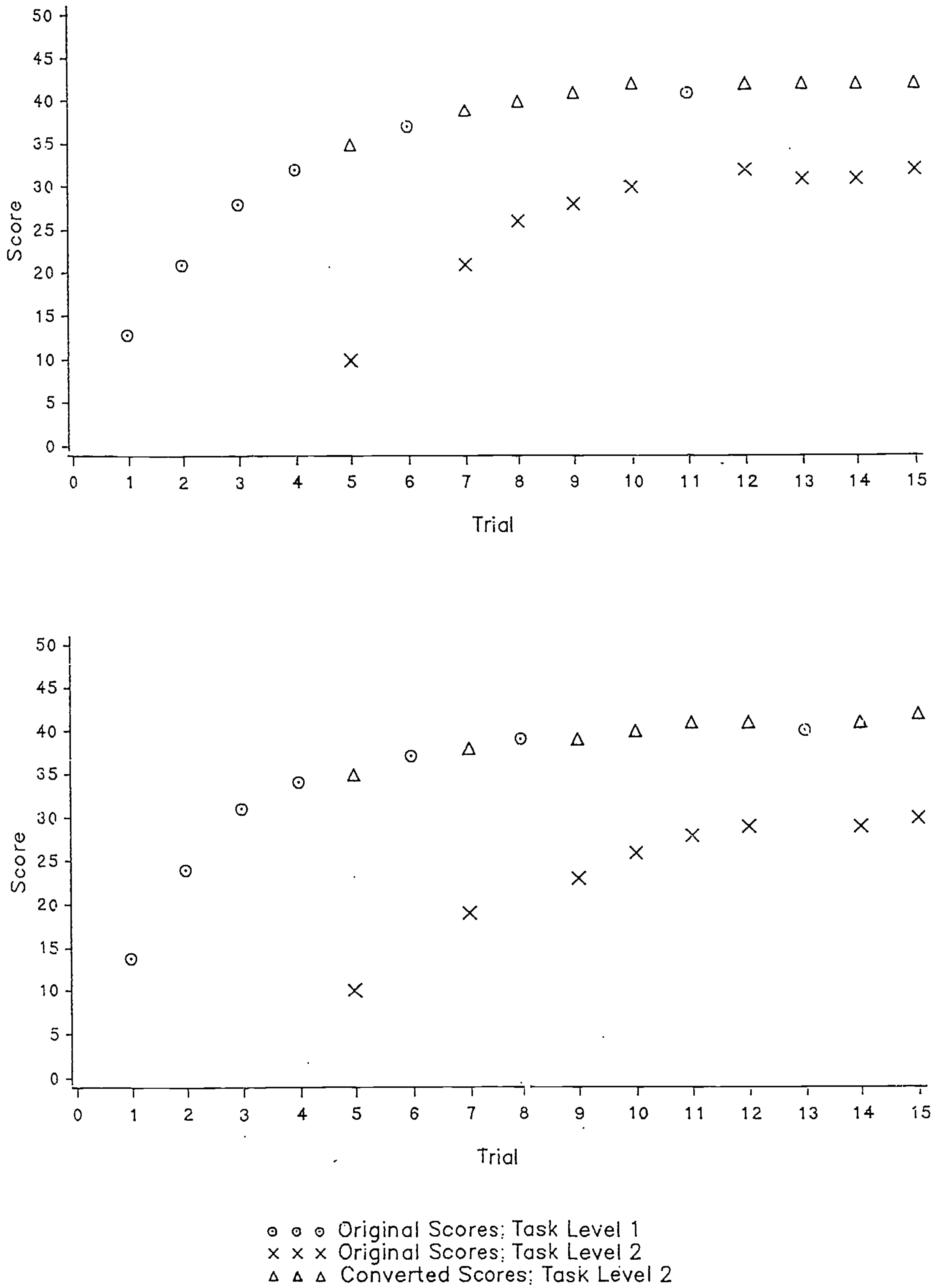


Figure 5. Examples of individual performances; simulation system, type 3.

decrement reduces as the trials progress. Other than the influence of the specific traits, the data for the illustrative examples are similar to the data for the preceding simulation system types.

Figure 6 gives the combined mean vectors. This time the vector using the observed task level 2 scores is much more ragged. However the conversion of the task level 2 scores produces a quite continuous function. Conversion coefficients summary for minimum RMSD at the succession of levels of differences is given in Table 8. There is not a large variability in these coefficients associated with level of differences. However, the least RMSD occurs at level of differences 3. A suggestion is to use the coefficients at that level of differences for which the RMSD is least. This time, there are no known theoretic conversion coefficients with

TABLE 8

CONVERSION COEFFICIENTS FOR MINIMUM RMSD

AT LEVELS OF DIFFERENCES 1 THROUGH 5

SIMULATION SYSTEM TYPE 3, RUN 1

Level of Differences	Coefficients a	b	RMSD
1	30.52	.320	3.51
2	231.46	.340	1.22
3	331.44	.337	.68
4	431.42	.339	1.05
5	531.37	.343	1.80

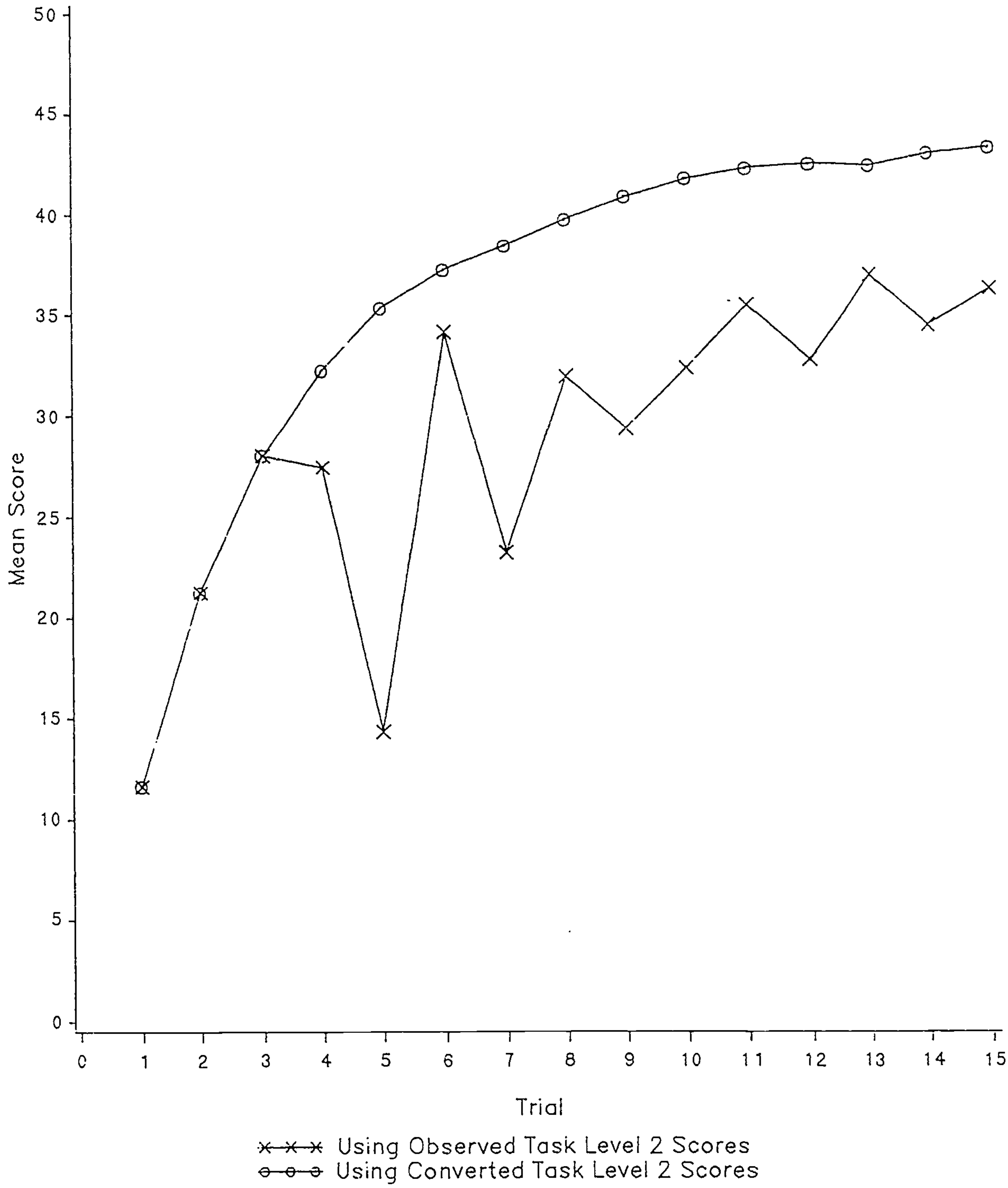


Figure 6. Mean scores for simulation system, type 3, run 1; conversion coefficients: a = 31.4, b = .34.

which to make a comparison. The effects of the dwindling influences of the specific traits has had an unknown extent of effect on the conversion.

Return to Figure 5. As in the previous simulation system types, the converted task level 2 scores tend, strongly, to form a continuous function which the task level 1 observed scores. The developed conversion appears to be satisfactory at the individual level.

Simulation System Type 4

The present type of simulation system was designed for an introductory look at another kind of complexity in which individuals did

TABLE 9

PARAMETERS FOR SIMULATION SYSTEM TYPE 4

RANDOM SCORE RANGES

Learning Curve Multiplier	35 to 55
Measurement Error	-1 to +1

TASK LEVEL 2 SCALING COEFFICIENTS

Additive	-10.0
Multiplicative	1.50

CUTTING SCORES

Task Level 1: 32.0 to Advance to Task Level 2

Task Level 2: 20.0 to Remain at Task Level 2

MAXIMUM NUMBER OF CONSECUTIVE TRIALS 4

SAMPLE SIZE 1000

not all follow the same learning function. In this type of simulation there are early learners and late learners with none in between. The mean function is followed by not one individual. Two basic learning functions are used with each individual being assigned to one or the other function on a random, 50-50 basis. An individual's true score on a trial was the basic learning function value for the trial times a multiplier for that individual. The range for these random multipliers is given in Table 9 which contains the parameters for this type of simulation system. At each trial for each individual a random measurement error was added to that individual's true score. Other parameters include the cutting scores for the two task levels, one to advance from task level 1 to task level 2 and the other to remain at task level 2. The program of review and look ahead was continued with a maximum number of consecutive trials at either task level. Sample size was 1000 .

Table 10 presents the two types of learning curves and the mean. Again, not one of the simulated individuals followed the mean curve. These basic learning curves are pictured in Figure 7. Type 1 learning curve is an early learning function while type 2 learning curve is a late learning function. The mean is half way between the two types of learning curves and represents not one individual. These learning curves are completely arbitrary and do not necessarily follow any mathematical function. They were written by the experimenter.

Figure 8 presents results for two simulated individuals. The upper illustration is for one of the early learners while the lower graph is for one of the late learners. Evidently, the scaling coefficients used for task level 2 made scores at this level on later trials higher than scores

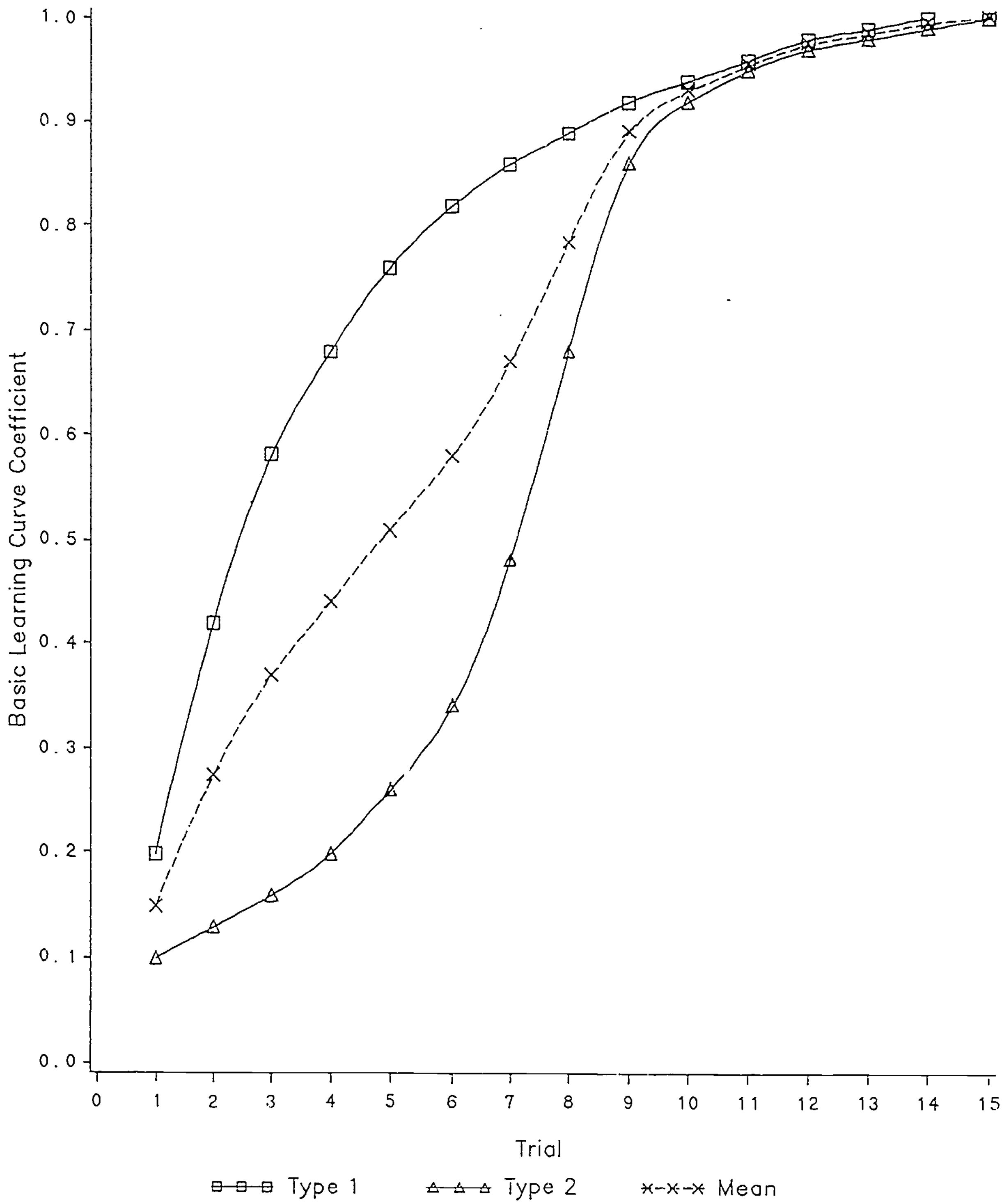
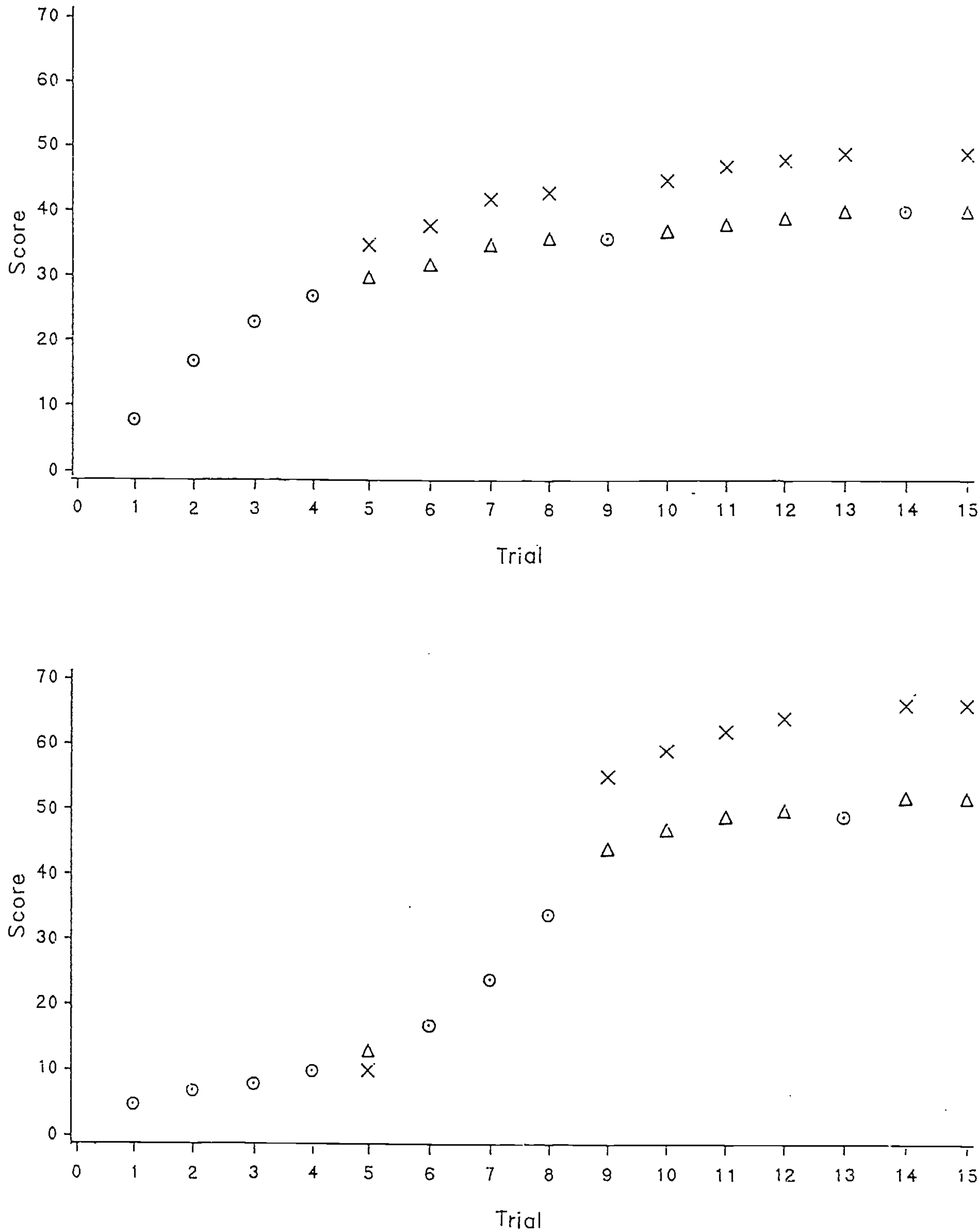


Figure 7. Basic learning curves for simulation system type 4.



⊙ ⊙ ⊙ Original Scores: Task Level 1
× × × Original Scores: Task Level 2
Δ Δ Δ Converted Scores: Task Level 2

Figure 8. Examples of individual performances: simulation system, type 4.

on task level 1. This may illustrate, also, the possibility of inaccurate judgments of task difficulty. The procedure for continuity should work properly with such inversions of task difficulties.

Figure 9 presents the two mean curves for run 1 of this type

TABLE 10

LEARNING CURVES

SIMULATION SYSTEM TYPE 4

	Basic Learning Curves			Mean Learning Curves	
Trial	1	2	Mean	Theoretic	Observed*
1	.200	.100	.150	6.7	6.7
2	.420	.130	.275	12.4	12.3
3	.580	.160	.370	16.7	16.6
4	.680	.200	.440	19.8	19.7
5	.760	.260	.510	23.0	22.7
6	.820	.340	.580	26.1	26.2
7	.860	.480	.670	30.2	30.3
8	.890	.680	.785	35.3	35.4
9	.920	.860	.890	40.0	40.0
10	.940	.920	.930	41.9	42.3
11	.960	.950	.955	43.0	43.4
12	.980	.970	.975	43.9	44.4
13	.990	.980	.985	44.3	44.7
14	1.000	.990	.995	44.8	44.8
15	1.000	1.000	1.000	45.0	45.5

*Task level 2 conversion coefficients: a = 5.99, b = .69.

BEST COPY AVAILABLE

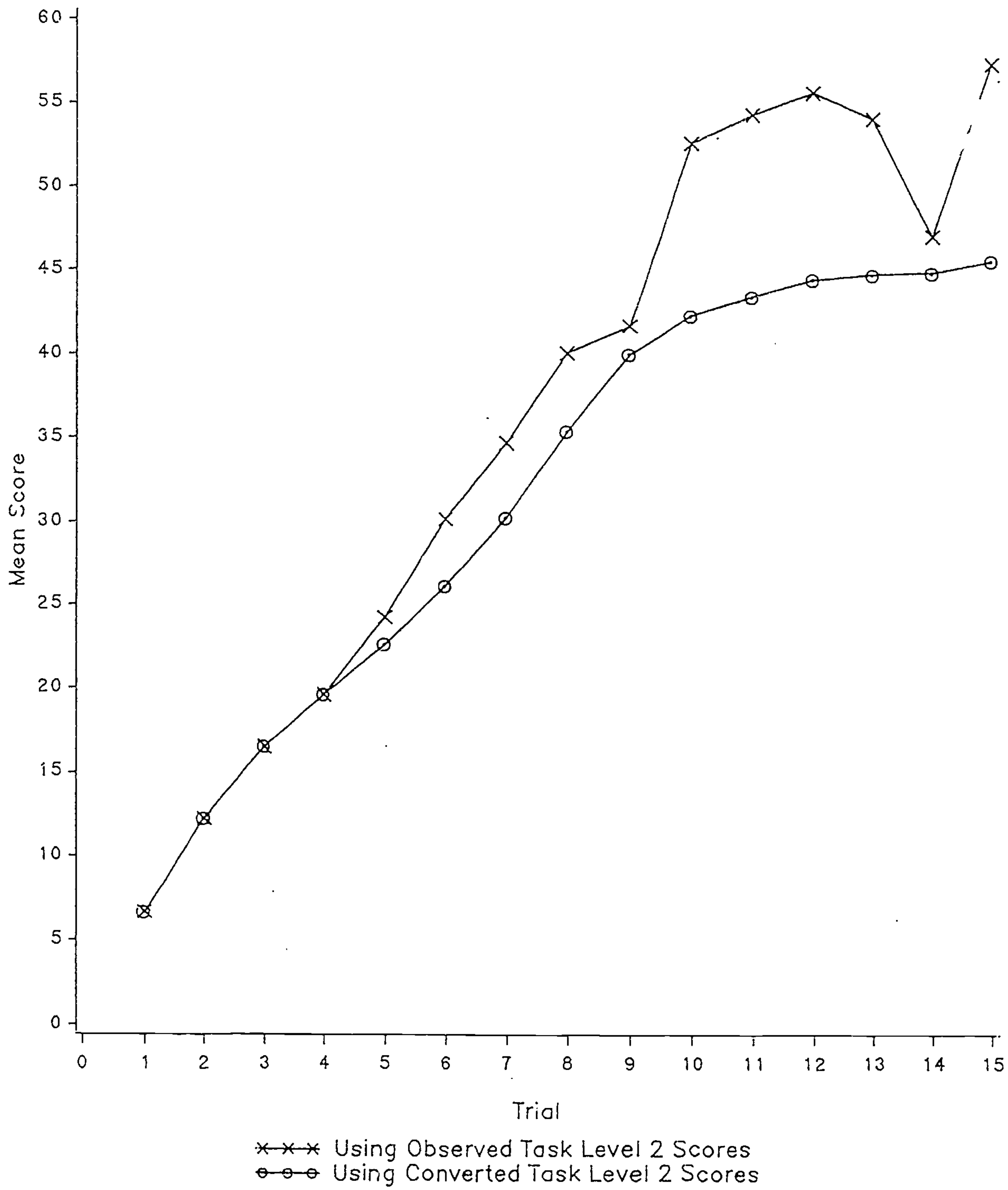


Figure 9. Mean scores for simulation system, type 4, run 1: conversion coefficients: a = 5.99, b = .69.

simulation system. As previously noted, the task level 2 obtained scores are higher than the task level 1 scores which makes the curve using the original task level 2 scores higher than the curve involving the converted task level 2 scores. While not one simulated individual followed the mean learning function, the combined mean vector using the converted task level 2 scores is quite continuous.

Table 11 contains the summary of conversion coefficients for minimum RMSD at the various levels of differences. After level of differences 1 the conversion coefficients have very little variability associated with level of differences. Since the least RMSD occurred at level of differences 3, the conversion coefficients at this level were used to convert the task level 2 scores. This conversion appears to work well for the simulated individuals as well as for the mean vector as is shown by the graphs in Figure 8. The mean combined score vector is used only as a

TABLE 11

CONVERSION COEFFICIENTS FOR MINIMUM RMSD

AT LEVELS OF DIFFERENCES 1 THROUGH 5

SIMULATION SYSTEM TYPE 4, RUN 1

Level of Differences	Coefficients a	b	RMSD
1	7.31	.641	3.26
2	6.16	.688	.98
3	5.99	.693	.92
4	5.93	.696	1.14
5	5.95	.696	1.66

device to look at the continuity principle. This use does not imply that the mean combined score vector represents any individual. A comparison is given in Table 10 between a theoretic mean curve and the observed mean curve. This theoretic mean curve is the mean individual multiplier times the mean basic curve.

Some preliminary Monte Carlo runs were made for simulation system type 4 as an initial investigation of stability of the developed conversion over sampling of individuals. The results appear to be very positive with very small sampling variability. More extensive Monte Carlo studies might be warranted with more complex simulation systems which more closely approach real observed phenomena.

Discussion

Calibration of measures from level to level in educational programs has been a long standing problem. From a lower level to a higher level the material and tasks may be greater both in difficulty and in complexity so that the assumption of unidimensionality of traits across levels of education is not warranted. Further, an educational program may have effects of increasing scores on latent traits (often conceived as factors in factor analytic theory). A principle of continuity of individual scores when individuals change from one level of task to another is proposed in this project as one method to ameliorate calibration problems in prolonged educational training programs. This principle is intended to include the concept of smoothness of function of performance on learning time. There should be no jumps or drops in this function at changes in task level nor should there be drastic changes in the slope of this function. There is no

BEST COPY AVAILABLE

requirement that the measures obtained at different levels of tasks be unidimensional in a continuation sense.

In our simulation system type 3, the two tasks had different specific traits in addition to a general trait. Scores on each of these traits increased with experience of the simulated individuals. In this example the influences of the specific traits were made to dwindle with practice so that effects of the general trait became dominant. In many psycho-motor type tasks, many experiments indicate the reverse to be more nearly true, the task specifics become dominant with practice. In cognitive development, quality educational programs should lead to enhancement of general traits with reduction of influences of task specifics. Application of the principle of continuity appeared to produce a satisfactory calibration of scores on task level 2 to be commensurate with observed scores on task level 1 in simulation system type 3.

The simulation system used in the present project involved a single general trait which carried over from one task level to another. Such a general trait could be a complex of more fundamental latent traits. Further, the dependencies of such a general trait could shift with experience of the individuals and there could be a shift in these dependencies from task level to task level. In addition scores of individuals on these more fundamental traits could be changing. Simulation systems to emulate these more complex situations could be written but would be quite long and involved. Such simulation systems were not attempted in the present project. Trials with such extended simulation systems could be informative and be accomplished in future projects.

The first three types of simulation systems used the negative exponential growth function in generation of individual data. However, the mean curves did not follow this function so that the success of the procedure was not linked directly to this growth function. In simulation system type 4 the mean curve was far from a negative exponential function. Apparently, the major requirement is that the combined mean curve be increasing, monotonic. This should not greatly restrict application of the principle of continuity.

During the development of the suggested procedure statistics other than the combined mean vector were explored. Similar to the combined mean vector, a vector of trial variances was considered. The task level 2 scores were converted by experimental conversion coefficients and the variances of scores for the trials were computed. Such vectors appeared not to yield results much different from those obtained with the combined mean vectors. Another area of investigation was the structure of generalized learning curves as per Tucker (1966). Again, no new information was apparent; however, this approach should be investigated for more complex situations.

Application of the principle of continuity probably would involve more than two levels of tasks. Procedures to extend the solution might be developed in future studies. One suggestion might go as described in the following. Make a tentative calibration of task level 2 scores to task level 1 scores. Then calibrate task level 3 scores to combined scores on task levels 1 and 2. If, there were only three task levels, then scores on each of the task levels could be calibrated with the combination of scores on the other two tasks. Such a procedure could be continued on a round robin program until overall stability was obtained. Some attention would be necessary to avoid scale drift during such iterations.

Only minimal attention was given in the present project to effects of sampling of individuals. A future project should consider this area of concern. A further question involves a possible requirement for complete data for each individual. How should incomplete data be handled? This question deserves further study to implement applications.

Experiments with real data would be highly desirable to provide information relevant to applications.

Reference

Tucker, L. R (1966). Learning theory and multivariate experiment: illustration by determination of generalized learning curves. In R. Cattell (Ed). *Handbook of multivariate experimental psychology* (pp. 476- 501). Chicago: Rand McNally.

ERIC

TECHNICAL NOTES

I The calibration problem

These notes will consider the calibration of scores on task level 2 to be comensurate with scores on task level 1. Observed scores are designated y_{mk} for an individual. A subscript for individual is not used for convenience. Task levels are designated as m = 1,2. Trials are designated k = 1, 2, ... ,n. Each individual has a score for each trial on either task level 1 or task level 2, never on both task levels. Task level 2 scores are to be converted to be comensurate with task level 1 scores. A linear conversion is used:

$$\tilde{y}_{2k} = a + by_{2k} \tag{TN-1}$$

where $\tilde{y}_{2k}$ is the converted score. Note that the task level 1 scores are used as observed so that:

$$\tilde{y}_{1k} = y_{1k} \quad . \tag{TN-2}$$

An individual combined score vector, $\tilde{\underline{y}}$ contains the $\tilde{y}_{1k}$ for trials when task level 1 was assigned to the individual and $\tilde{y}_{2k}$ for trials when task level 2 was assigned. The calibration problem is to determine the conversion coefficients a and b so that each and every combined score vector is as smooth and continuous as possible. Of course, there may be occasions when such a solution leads to unsatisfactory results.

II <u>General plans for generation of simulated data</u>

Generated scores for a simulated individual are indicated as follows.

Skill score x_{mk} for task level $m = 1, 2$ and trial $k = 1,2,\ldots,n$.

True task score t_{mk}:

$$t_{mk} = f_m + g_m x_{mk} \qquad \text{(TN-3)}$$

where f_m and g_m are scaling coefficients.

Observed task score y_{mk}:

$$y_{mk} = t_{mk} + e_{mk} \qquad \text{(TN-4)}$$

where e_{mk} is a random error of measurement.

In the real world, skill score x_{mk} is the value of a latent variable for the individual. In simulations, x_{mk} is computed by procedures defined by the simulation system and depends upon random parameters for the simulated individual.

Each individual is assigned to work at task levels according to the following plan.

Start at task level 1 and continue until the individual's observed score exceeds a cutting score y_{1c} or is advanced for experience in work at task level 2 after a given number of trials at task level 1.

When working at task level 2, remain at this level unless the individual's score is less than a cutting score y_{2c} or the individual has worked at this level for a given number of trials and was scheduled for a review at task level 1.

Parameters in the simulation systems follow.

y_{1c} : cutting score to advance from task level 1 to task level 2.

y_{2c} : cutting score to remain at task level 2.

Maximum number of trials to remain at a given task level.

To facilitate subsequent computations a score vector $\underline{v}$ is set up for each individual with 4n elements composed of the following 4 sections. Each section has n elements.

$\underline{\delta}_1$: dummy variables for the trials with score = 1 if the individual was assigned to task level 1, otherwise = 0.

$\underline{y}_1$: scores on task level 1 if the individual was assigned to that level, otherwise = 0.

$\underline{\delta}_2$: dummy variables for the trials with score = 1 if the individual was assigned to task level 2; otherwise = 0.

$\underline{y}_2$: scores on task level 2 if the individual was assigned to that level of task; otherwise = 0.

A mean vector M_v and mean product matrix P_{vv} are accumulated over the individuals. The mean vector and mean product matrix have sections corresponding to the sections of vector $\underline{v}$.

III <u>Statistics for combined score vectors</u>

A combined score vector $\tilde{\underline{y}}$ was defined in section I with converted scores defined in equations (TN-1) and (TN-2). For any given pair of calibration coefficients a and b, a combined score vector may be obtained using the score vector $\underline{v}$ defined in section II and a matrix W (n x 4n) defined as follows:

$$W = [0,\ I,\ A,\ B] \qquad \text{(TN-5)}$$

where 0 is an (n x n) null matrix; I is an (n x n) identity matrix; A is an (n x n) scalar matrix containing coefficient a in the diagonal; and B is an (n x n) scalar matrix containing coefficient b in the diagonal. Since vector sections $\underline{\delta}_1$ and $\underline{y}_1$ have zero entries when the individual was assigned to task level 2 while sections $\underline{\delta}_2$ and $\underline{y}_2$ have zero entries when the individual was assigned to task level 1, the operations of equations (TN-1) and (TN-2) may be combined to yield the combined score vector $\tilde{\underline{y}}$ by:

$$\tilde{\underline{y}} = \underline{v}W' \qquad \text{(TN-6)}$$

where $\tilde{\underline{y}}$ and $\underline{v}$ are row vectors.

Since equation (TN-6) is a linear transformation, the combined score mean vector $M_{\tilde{y}}$ may be obtained from the mean vector M_v by:

$$M_{\tilde{y}} = M_v W' \quad . \qquad \text{(TN-7)}$$

Note that $M_{\tilde{y}}$ and M_v are row vectors. Also due to the linear transformation, the mean product matrix for the combined scores may be obtained from the mean product matrix P_{vv} by:

$$P_{\tilde{y}\tilde{y}} = WP_{vv}W' \quad . \tag{TN-8}$$

The variance of the combined scores for each trial may be obtained from the corresponding diagonal entry in $P_{\tilde{y}\tilde{y}}$ and the corresponding mean in $M_{\tilde{y}}$. These variances may be recorded in a vector termed VAR.

ERIC

IV Testing for smoothness by vector differences

The present discussion will focus on the relation of the combined score means to the trials. A similar development applies to the trial variances in vector VAR . Combined score means, $\bar{y}_k$, are taken to be measured on a continuous scale while the trials are taken to be discrete points on a time-like dimension. Combined score means form a dependent variable while trials form an independent variable. In cases when both the independent variable and the dependent variable are continous, derivatives are indicative of properties of the functional relation with the first derivative being the rate of change, the second derivative being the acceleration, third and higher derivatives being indicators of irregularities such as jerks or jumps. In the case when the independent variable exists only at equally spaced points, differences between values of the dependent variable take the place of the derivatives.

Table 3 presents two examples of combined mean vectors for simulation system type 1. Each of these combined mean vectors is for a selected pair of calibration coefficients. The vector differences are computed out to the fifth level of differences. These differences may be defined by the following equations. Let level of difference be indicated by $j = 1,2,\ldots$. Let d_{jk} be the jth level difference at trial k . The first level differences are obtained from the combined mean values:

$$d_{1k} = \bar{y}_k - \bar{y}_{(k-1)} \quad \text{for } k = 2,3,\ldots,n. \qquad \text{(TN-9)}$$

For second level of differences and higher, the values are obtained from the preceding levels of differences. For $j = 2, 3, \ldots$:

$$d_{jk} = d_{(j-1)k} - d_{(j-1)(k-1)} \quad \text{for} \quad k = (j+1), (j+2), \ldots, n. \tag{TN-10}$$

The bottom row of each section of Table 3 gives the RMSD statistic for the differences at each level of difference. These are the root mean square statistics for the differences. Note that the number of differences reduces with the level of differences, this number being $(n - j)$. The RMSD statistic is used as a general index of the magnitudes of the differences. Comparison of the differences between the two pairs of calibration coefficients indicates a general reduction in the magnitudes of the differences after level of differences 1.

V Determination of calibration coefficients for minimum RMSD

For any given level of differences the RMSD statistic can be considered as a dependent variable with a bivariate surface on the calibration coefficients a and b as independent variables. Table TN.1 presents values of RMSD for selected values of a and b for level 5 of results for simulation system type 1. For each value of b there is a minimum RMSD in the range tabled of a. This might be expected since a is an additive constant which moves the converted task level 2 scores up and down. For a fixed value of b, there should be a best such translation of the task level 2 scores. Deviation of the translation from this best value should produce some jerk or drop in the combined score vector and lead to an increase in the RMSD statistic for higher level differences. Note that there are also minima in the rows for middle values of a. This surface appears to be sufficiently regular to permit a solution for values of a and b to yield a minimum RMSD.

Inspection of the RMSD surface indicated the possible use of a parabolic interpolation procedure to determine for a given value of b the value of a for a minimum RMSD. This procedure is incorporated in subroutine VALA which may be inspected for details. The general idea is to use three points on the function of RMSD on a. Fit a parabola to these three points and solve for the location of the value of a for a minimum of the parabola. Then, the value of RMSD is computed for this new value of a. From these results a revised three points are established and a new solution obtained. This procedure is continued until a converged solution is obtained.

TABLE TN.1

ILLUSTRATION OF DEPENDENCE OF RMSD STATISTIC

ON CALIBRATION COEFFICIENTS

Simulation System Type 1; Combined Mean Vector

Level of Differences = 5

Value of a	Value of b .4	.5	.6	.7	.8	.9	1.0	1.1	1.2
0	12.9	11.7	10.4	9.2	7.9	6.7	5.4	4.2	3.0
4	10.9	9.7	8.4	7.2	5.9	4.7	3.5	2.2	1.1
8	9.0	7.7	6.5	5.2	4.0	2.7	1.5	.4	1.2
12	7.0	5.7	4.5	3.2	2.0	.7	.6	1.8	3.1
16	5.0	3.7	2.5	1.2	.1	1.3	2.5	3.8	5.0
20	3.0	1.8	.6	.8	2.0	3.3	4.5	5.8	7.0
24	1.1	.5	1.5	2.8	4.0	5.6	.5	7.8	9.0
28	1.1	2.3	3.5	4.8	6.0	7.3	8.5	9.8	11.0
32	3.0	4.3	5.5	6.7	8.0	9.2	10.5	11.7	13.0

For some details of this procedure, consider the general parabola:

$$y = a + bx + cx^2 \quad . \tag{TN-11}$$

Solution for the value of x for the optimum y yields

$$x_n = \frac{[(y_3 - y_2)(x_2^2 - x_1^2) + (y_1 - y_2)(x_3^2 - x_2^2)]}{2[(y_3 - y_2)(x_2 - x_1) + (y_1 - y_2)(x_3 - x_2)]} \tag{TN-12}$$

where x_n is the value of x at optimum y. Whether the optimum is a maximum or a minimum depends upon the configuration of the three points (x_1,y_1), (x_2,y_2) and (x_3,y_3). In the present application a minimum would be normal from the configuration of values in each column of the RMSD surface.

The three points are chosen such that the x's are in ascending value with y_2 being less than both y_1 or y_3. With an initial value of x_2 there is a search for such a series of points. When such a series of points has been established, a value of x_n (x new) is obtained by equation (TN-12). Given x_n, the actual value of y_n is computed, not from the parabola but from the observations. From the now four points a selection is made of the three points is made such that the interval from x_1 to x_2 is least and y_2 is less than the the other two y's. This procedure may be continued until the difference between x_n and x_2 is less than some set tiny value.

The parabolic interpolation scheme was used to determine a for given values of b to obtain minimum RMSD. Determination of b involved the same procedure. This time the dependent variable y was the minimum value of RMSD conditional on the value of b. Subroutine VALAB provides these computations resulting in the solution values for a and b.

A solution for a and b may be made at each of several levels of differences and the value of RMSD computed. It appears that the level of differences should be selected for which the minimum RMSD is least.

ERIC

V Exploration for structure of combined mean product matrix

Tucker's development of generalized learning curves (1966) forms the basis for exploration for structure of combined mean product matrices. The major idea is that optimum calibration coefficients should lead to least complex structure of the mean product matrix. Inappropriate calibration of task level 2 scores should induce complexities in this structure.

The major computational technique uses an eigen solution of each combined mean product matrix. The series of eigenvalues may be inspected for indications that there is a small number of relevant dimensions involved in the matrix. See Table TN.2 for examples of series of eigenvalues of combined mean product matrices obtained for simulation system type 2, run 1, using two pairs of calibration coefficients. On the left is the series when the raw task level 2 scores are combined with the raw task level 1 scores. On the right is the series using the calibration coefficients obtained for minimum RMSD of the combined mean vector, level 5 differences. There are several more moderate-sized eigenvalues in the left series than in the right series. This is indicative of a greater complexity of the combined mean product matrix using the raw task level 2 scores than for the combined mean product matrix using the calibration coefficients obtained from the combined mean vector. Inspection of such series of eigenvalues is one means for evaluation of the complexity of combined mean product matrices.

Complexity of structure is indicated also in the eigenvectors of a combined mean product matrix. Table TN.3 lists the first three eigenvectors for the two combined mean product matrices discussed in the preceding paragraph. These eigenvectors are graphed in Figure TN.1. The differences in complexity are more apparent in the figure. Note the considerable irregularities of the series of coefficients of the eigenvectors for the combined mean product matrix using raw task level 2 scores. In contrast, the eigenvector coordinates form smooth series for the combined mean product matrix using the optimum calibration coefficients for combined mean vectors. Thus, the eigenvectors provide further indications as to complexity of combined mean product matrices.

Ways to use the complexity of combined mean product matrices have not been studied to any great extent. At a minimum, the complexity of a resulting combined mean product matrix should be checked when calibration coefficients have been determined by other procedures. As yet, no index of complexity has been developed which could be used in determining optimum calibration coefficients. This could be an area for further study.

TABLE TN.2

ANALYSIS OF COMBINED MEAN PRODUCT MATRIX

EIGENVALUES

Simulation System Type 2, Run 1

Calibration Coefficients

a = .00 b = 1.000

Dimension	Value	Difference
1	267.7	
		191.4
2	76.3	
		28.2
3	48.1	
		17.6
4	30.5	
		9.5
5	21.0	
		14.7
6	6.3	
		1.2
7	5.1	
		1.6
8	3.5	
		1.7
9	1.8	
		.1
10	1.7	
		.7
11	1.0	
		.1
12	.9	
		.5
13	.4	
		.1
14	.3	
		.1
15	.2	

Calibration Coefficients

a = 16.11 b = .795

Dimension	Value	Difference
1	253.7	
		193.3
2	60.4	
		58.7
3	1.7	
		1.5
4	.2	
		.0
5	.2	
		.0
6	.2	
		.1
7	.1	
		.0
8	.1	
		.0
9	.1	
		.0
10	.1	
		.0
11	.1	
		.0
12	.1	
		.0
13	.1	
		.0
14	.1	
		.0
15	.1	

TABLE TN.3

ANALYSIS OF COMBINED MEAN PRODUCT MATRIX

EIGENVECTORS

Simulation System Type 2, Run 1

Calibration Coefficients

a = .00 b = 1.000

	Dimension		
Trial	1	2	3
1	.16	-.29	.24
2	.25	-.36	.30
3	.25	-.29	.19
4	.11	.14	.36
5	.36	-.22	.23
6	.10	.02	-.18
7	.28	.02	.05
8	.36	-.29	-.36
9	.28	.36	.33
10	.27	.12	-.1
11	.16	.26	-.22
12	.29	.19	-.07
13	.35	-.11	-.45
14	.25	.48	.21
15	.23	.25	-.24

Calibration Coefficients

a = 16.11 b = .795

	Dimension		
Trial	1	2	3
1	.20	-.34	.51
2	.29	-.40	.35
3	.32	-.34	.06
4	.33	-.24	-.15
5	.32	-.14	-.26
6	.30	-.04	-.31
7	.29	.04	-.29
8	.27	.11	-.24
9	.25	.17	-.16
10	.23	.21	-.05
11	.22	.25	.04
12	.21	.28	.13
13	.20	.30	.23
14	.19	.32	.28
15	.19	.33	.35

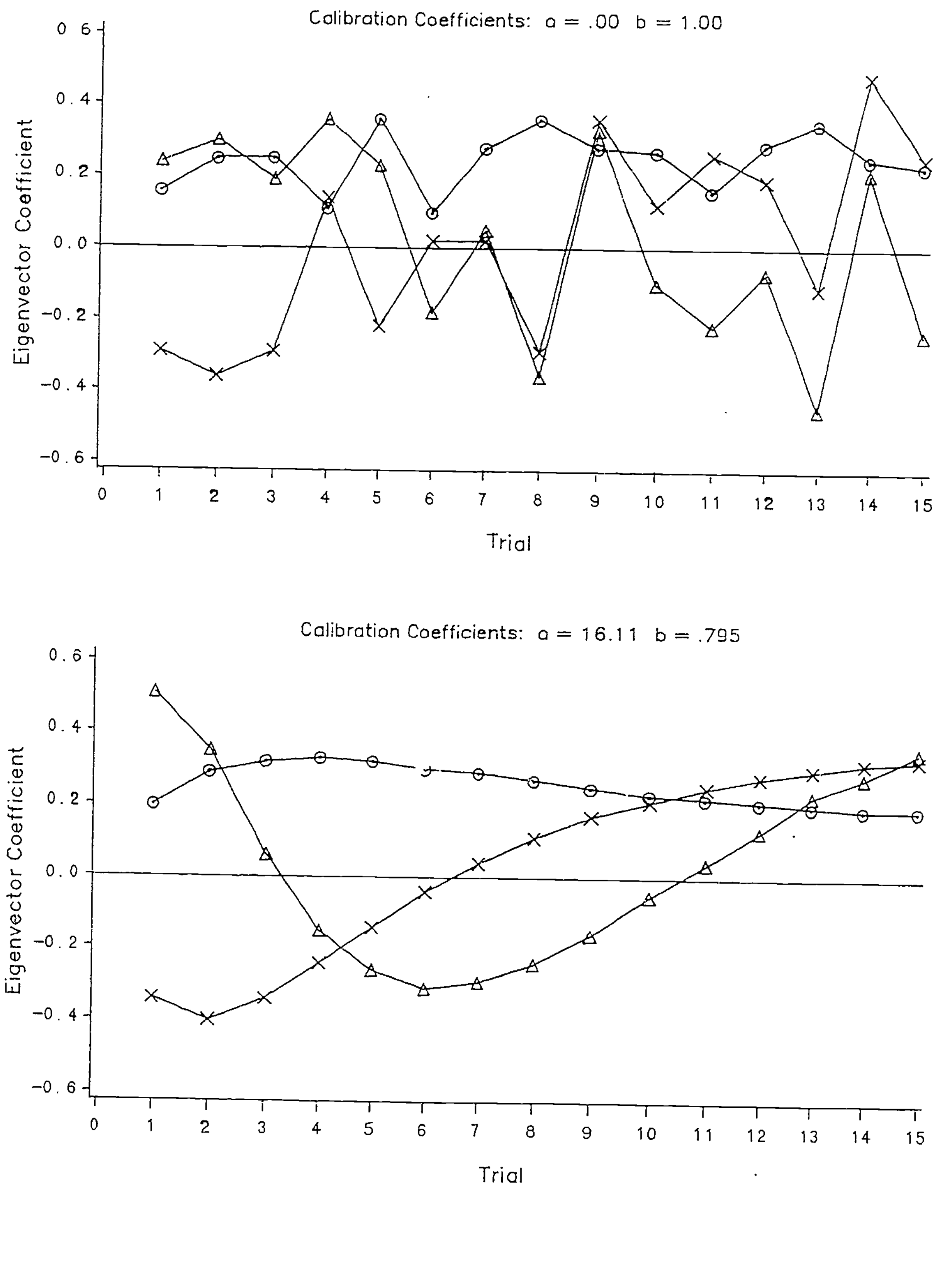


Figure TN.1. Graphs of eigenvectors of combined mean product matrices.

VII Procedures for generation of simulated scores

Random scores with a triangular distribution are used in a number of places in the generation of random scores. The reason for this choice is that in almost all situations there is a restriction that scores be positive. It is convenient to use scores with defined ranges. Let s be a generalized random score which is to be distributed triangularly in the range from s_ℓ to s_u. Two random p's, p_1 and p_2, are drawn from a rectangular distribution with a range 0 - 1 . Then:

$$s = a_d + b_d(p_1 + p_2) \qquad \text{(TN-13)}$$

where

$$a_d = s_\ell \quad ; \qquad \text{(TN-14)}$$

$$b_d = \frac{1}{2}(s_u - s_\ell) \qquad \text{(TN-15)}$$

A negative exponential learning function is used in simulation system types 1, 2, and 3. Let x_k be a general skill score at trial k. The negative exponential learning function is:

$$x_k = u - e^{c(d - k)} \qquad \text{(TN-16)}$$

where u is the asymptotic value of x , c and d are individual parameters. A necessary inequality for this learning function is that:

$$u > x_1 > 0 \qquad \text{(TN-17)}$$

where x_1 is the skill score of the individual on the first trial. For each individual a first trial score and an asymptotic score are drawn randomly such that the inequality of equation (TN-17) holds. With the triangularly distributed scores, the ranges for x_1 and u should not overlap, with the range for u being higher than the range for x_1. Parameters c and d are determined by:

$$c = \ln(u) - \ln(u - x_1) \qquad \text{(TN-18)}$$

$$d = [\ln(u)]/c \qquad \text{(TN-19)}$$

Simulation system type 1 is quite simple involving a single skill score for each individual which increases with trials of either task level 1 or task level 2 following a negative exponential learning function. Scaling coefficients f_1 and g_1 are set to 0 and 1 for task level 1. Scaling coefficients f_2 and g_2 are parameters for the particular run. No measurement error was added to the true task scores. The program of review and look ahead was not implemented.

Simulation system type 2 is very similar to simulation system type 1. In this second system error of measurement is added to the true task scores and the program of review and look ahead is implemented.

Simulation system type 3 introduces a complexity in generation of scores. In addition to a general skill there is a specific skill for each task level so that the skill score for each task is a combination of the general skill and the specific skill for that task. The general skill increases with every trial while the specific skill for a task level increases only with practice at that task level. This introduces some programing complexities due to the need to keep track of the number of trials for each level of task.

Let h_m be the trial number for task level m. Then:

$$k = h_1 + h_2 \qquad \text{(TN-20)}$$

Let z_{gk} be the general skill score at trial k and z_{smh} be the specific skill score for task level m at trial h_m. A linear composite of these scores produces the task skill score.

$$x_{mk} = w_{gmh}z_{gk} + w_{smh}z_{smh} \qquad \text{(TN-21)}$$

where w_{gmh} is the weight for the general skill at trial h_m for task level m and w_{smh} is the weight . .: specific skill for task level m at trial h_m for that task level. The general skill score and each of the specific skill scores increase by the negative exponential learning function using the appropriate number of trials, k for the general skill score and h_m for each of the specific skill scores. There are separate individual parameters for the general skills learning function and the specific skills learning function. The weights are adjusted for each task level so that the dependence on the specific skill reduces and the weight for the general skill increases as practice continues:

$$w_{gmh} = h_m/(h_m + 1) \qquad \text{(TN-22)}$$

$$w_{smh} = 1/(h_m + 1) \quad . \tag{TN-23}$$

These weights sum to unity.

The scaling parameters, addition of errors of measurement, and program of review and look ahead are as in simulation system type 2.

Simulation system type 4 is a variant in which different individuals follow different learning functions. The negative exponential learning curve is not used. Instead, a table of learning functions is read in and a choice is made for each individual as to which learning function is to be followed. Let matrix Z contain the learning functions with a row for each trial and a column for each learning function. For each individual there are two parameters: j, equal to the learning function to be used, and a multiplier, w. Parameter j is a random digit. (Equal probabilities for 1 and 2 were used). Parameter w is a random value in a defined range. The trial skill score is:

$$x_{mk} = z_{jk}w \tag{TN-24}$$

In all other aspects, simulation system type 4 is like simulation system type 2.